Advances in Water Resources 112 (2018) 147–159

Contents lists available at ScienceDirect

Advances in Water Resources

journal homepage: www.elsevier.com/locate/advwatres

Regional-scale analysis of extreme precipitation from short and fragmented records

Andrea Libertino[*,a], Paola Allamano[b], Francesco Laio[a], Pierluigi Claps[a]

[a] Dipartimento di Ingegneria dell'Ambiente, del Territorio e delle Infrastrutture, Politecnico di Torino, Corso Duca degli Abruzzi, 24, Torino 10129, Italy
[b] Waterview SRL, Corso Castelfidardo, 30/A, Torino 10129, Italy

ARTICLE INFO

Keywords:
Regional analysis
Annual maxima
Fragmented records
Rainfall extremes
L-moments

MSC:
86A05

ABSTRACT

Rain gauge is the oldest and most accurate instrument for rainfall measurement, able to provide long series of reliable data. However, rain gauge records are often plagued by gaps, spatio-temporal discontinuities and inhomogeneities that could affect their suitability for a statistical assessment of the characteristics of extreme rainfall. Furthermore, the need to discard the shorter series for obtaining robust estimates leads to ignore a significant amount of information which can be essential, especially when large return periods estimates are sought. This work describes a robust statistical framework for dealing with uneven and fragmented rainfall records on a regional spatial domain. The proposed technique, named "patched kriging" allows one to exploit all the information available from the recorded series, independently of their length, to provide extreme rainfall estimates in ungauged areas. The methodology involves the sequential application of the ordinary kriging equations, producing a homogeneous dataset of synthetic series with uniform lengths. In this way, the errors inherent to any regional statistical estimation can be easily represented in the spatial domain and, possibly, corrected. Furthermore, the homogeneity of the obtained series, provides robustness toward local artefacts during the parameter-estimation phase. The application to a case study in the north-western Italy demonstrates the potential of the methodology and provides a significant base for discussing its advantages over previous techniques.

1. Introduction

Probabilistic modelling of extreme rainfall has a crucial role in flood risk estimation and consequently in the design and management of flood protection projects (Koutsoyiannis, 2007). The first attempts to establish a mathematical relation between intensity and frequency of rainfall goes back to as early as 1932 (Bernard, 1932). Since then, many studies (e.g., Svensson and Jones (2010)) have been carried out, aimed at providing the rainfall depths for different return periods and durations. Complete overviews on the different approaches adopted from several countries around the globe can be found, e.g, in Castellarin et al. (2012); Szolgay et al. (2009).

Intensity-Duration-Frequency (*IDF*) and Depth-Duration-Frequency (*DDF*) curves are commonly adopted in water resources engineering for both planning, designing and operating of water resource projects and for land and people protection purposes (Koutsoyiannis et al., 1998). These curves are usually developed considering the historical records for different durations and adopting the index-rainfall method, in which the quantile of the extreme rainfall comes as the product of an "index value" (i.e., usually the mean) and a growth curve (i.e., the non-dimensional inverse of the frequency distribution $F(x)$).

Two approaches are commonly adopted for fitting a probability distribution to the series of maxima: (i) the "block" method, that consists in selecting the maximum rainfall occurring over a fixed period (usually 1 year) and (ii) the "peak-over-threshold" method, in which all the rainfall data exceeding some pre-specified threshold are considered (Coles, 2001). The method (i) is widely adopted in Italy for design rainfall estimation, and a large dataset of annual maxima for duration 1-3-6-12-24 h is available, which dates back to the early twentieth century.

Due to the significant developments of the theory of extreme value in the last two decades (Coles, 2001; Reiss and Thomas, 2001) the methodologies for rainfall frequency analysis are nowadays quite established and robust, both at the single-station and at the regional scale. However, the correct reproduction of complex hydro-meteorological processes requires not only long, but also serially complete and reliable observations (Koutsoyiannis, 2004; Pappas et al., 2014) from a dense and spatially uniform monitoring network. A non-uniform and non-

* Corresponding author.
E-mail address: andrea.libertino@polito.it (A. Libertino).

https://doi.org/10.1016/j.advwatres.2017.12.015
Received 31 July 2017; Received in revised form 15 December 2017; Accepted 15 December 2017
Available online 16 December 2017

continuous dataset can prevent a reliable application of the aforementioned methodologies at the regional scale leading to inconsistencies.

It is thus evident that, despite the existence of established rainfall frequency analysis techniques, operational and methodological problems concerning their applications still arise.

Rainfall time series are often plagued with missing values creating sporadic and/or continuous gaps in their records. The fragmented behaviour traces back to the activation and dismissal of rain gauges, attributable to station relocation, service interruptions, replacement/renewal of the sensor, changes in the ownership of the station, etc. The characteristics of the stations (location and elevation, type of sensor, etc.) may also change before and after the interruptions, with consequent problems in attributing the data to a unique homogeneous sample. Despite these problems are quite common, even in developed countries, many practical applications and statistical methodologies have little or no tolerance to missing values (Pappas et al., 2014; Teegavarapu and Nayak, 2017)

The treatment of gaps in the records or relocation of rain gauges, especially when dealing with large databases, requires the set-up of specifically-conceived methodologies aimed at bypassing or reconciling the inconsistencies (Acquaotta et al., 2009). Two approaches can be adopted for dealing with non-uniform sets of records: (i) a precautionary approach, that consist in assuming a minimum acceptable threshold of record length and discard the series shorter than the threshold and (ii) a preservative approach, focused on the identification of methodologies aimed at extracting all the available information even from the shorter records. While, on the one hand the approach (i) can discard important information hidden in the shorter records, affecting the results of the regional rainfall frequency analysis, the approach (ii) turns out to be complex, computationally demanding, and can lead to errors when based on non-robust assumptions (Teegavarapu and Nayak, 2017).

A number of procedures for recovering information from short records can be found in the literature. Various authors propose the adoption of interpolation techniques along the time-axis, to estimate the missing data of environmental series (linear or logistic regression, polynomial or spline interpolation, inverse distance weighting, ordinary kriging, etc. - see, e.g., Koutsoyiannis and Langousis (2011); Maidment et al. (1992)). The statistical techniques available include also artificial neural networks and nearest neighbours (Elshorbagy et al., 2002; 2000), approaches based on Kalman filters (Alavi et al., 2006), non-linear mathematical programming (Teegavarapu, 2012b) and normal-ratio and inverse distance weighting methods (Teegavarapu and Chandramouli, 2005).

In Pappas et al. (2014) it is argued that the complexity and the computational burden associated with these techniques often make them unsuitable for an application over large scales. This usually leads to the adoption of conceptually over-simplified approaches (e.g., filling the gaps with fixed values, often corresponding to the sample average of the series) not adequate to represent the complexity of the phenomena. The authors propose a simple method based on the analysis of the auto-correlation structure of the series, amenable for a quick filling of sporadic gaps. However, the technique is viable if the percentage of missing values in the time series is limited. When the gaps are frequent and systematic (e.g., in developing countries Clarke et al. (2009)) and when data show low auto-correlation in time, this approach is not effective.

Even when long uninterrupted rainfall records are available, an *IDF* relation is basically valid only at the point where it is estimated. Rain gauges are generally not evenly distributed in space, and they allow only for a point estimation of the parameters of the rainfall distribution. To extend estimates to ungauged locations, rainfall data are usually interpolated, either by considering the distribution parameters estimated at the station location (e.g., Ashraf et al. (1997); Myers (1994)), or by estimating the *IDF*s after pooling the available data within homogeneous areas defined by geographical boundaries, or centred around a location of interest (see, e.g., Hosking and Wallis (1997)). In the presence of data scarcity, some recent studies also propose to include external sources of data (e.g., remote sensing data Qamar et al. (2017)) in the procedure. Regional techniques for rainfall frequency analysis actually build representative growth curves from larger samples resulting from pooling. On the other hand, the use of a regional frequency curve is suitable only when the spatial dependence is weak enough to enable transferring information to a site of interest from the surrounding gauged sites (Buishand, 1991). When spatial dependence is significant, as in the presence of high discontinuity in the rainfall distribution, or due to different climatic and orographic conditions, different approaches should be preferred. For instance, Uboldi et al. (2014) propose a statistical approach that involves the adoption of a bootstrap algorithm aimed at providing complete annual maxima series at each location, taking into account all data observed at surrounding stations with decreasing importance when distance increases. This kind of approach allows one to overcome the problem of data filling, but the bootstrap procedure produces results that deviate significantly from the sample spatial distribution, ignoring the existence of long and reliable records at some locations.

In this work, a simple approach able to provide a set of complete series of rainfall data for each location of the domain under analysis is proposed. The methodology, described in section 3.1, is summarized in figure 1. It is based on the sequential application of the ordinary kriging equation to the values recorded annually in the region of interest. The so-called “patched kriging” procedure preserves the spatio-temporal information of the annual maxima recorded by the monitoring network, “patching” them together, i.e., considering each record just like a point in the (x, y, t) space (where x and y are the planimetric coordinates and t is the time).

From an operational point of view, this methodology has a low computational cost and does not require to work with stationary or significantly auto-correlated data, as it does not involve any interpolation along the time-axis. This feature proves to be particularly effective when dealing with frequent rain gauge relocations, allowing on the one hand to maximize the usable information at gauged sites, and on the other to extend the analysis to the ungauged ones.

2. Data and case study

The region considered for the demonstration of the proposed methodology refers to the Piemonte region, an area of about 30,000 km^2 in the North-Western part of Italy, shown in Fig. 2a. The area is characterized by a very heterogeneous orography, flat or hilly in the centre, surrounded by the Alps in the North-West and by the Ligurian Apennines in the South, with the minimum elevations of the order of a few tens of meters a.s.l. and the maximum ones exceeding 4000 m a.s.l. Several regional-scale hydrological analyses have been performed with a focus on this area (e.g. Ganora et al. (2013); Laio et al. (2011); Qamar et al. (2015)); in all cases, the availability of accurate extreme-rainfall statistics is an essential prerequisite for obtaining consistent results.

A dataset of annual maximum rainfall depths over duration intervals of 1, 3, 6, 12 and 24 h from 1928 to 2010 has been assembled for this analysis. The data before the ’90s were collected from the publications of the National Bureau for Hydro-Meteorological Monitoring (*SIMN*). After 1987 the network was gradually taken over by the Regional Environmental Agency (*ARPA* Piemonte) that removed, substituted or relocated some of the stations. Gauge data from neighbouring regions has also been considered to limit the edge effects. Overall, nearly 500 gauging stations have worked for at least one year in the considered period.

Annual maximum values have been extracted from the original rainfall series by the competent authorities using sliding time windows (van Montfort, 1990; Papalexiou et al., 2016). The original series have a resolution in time varying from 1 h for the oldest stations to 5 min. for